

基于骨骼和深度信息的手势识别的研究与应用

吴彩芳¹, 谢 钧¹, 俞 璐²

(1. 解放军理工大学 指挥信息系统学院, 江苏 南京 210007;
2. 解放军理工大学 通信工程学院, 江苏 南京 210007)

摘 要:文中研究了基于 Kinect 的手势识别技术,设计并实现了一个功能完善、性能优越的小型手势交互系统。首先结合 Kinect 获取的人体骨骼信息和深度信息,实现了手部的追踪和提取,并且实验效果不受实验背景、光线、实验者的肤色和服装的影响。然后根据初步获取的手型二值图噪声分布特点,提出一种过滤小规模连通分量像素点的方法对二值图进行去噪。最后,分别以手型二值图 Hu 矩和手型轮廓二值图 Hu 矩为特征,使用 SVM 分类器进行训练和识别。实验结果表明,同手型二值图的 Hu 矩相比,以手型轮廓二值图的 Hu 矩作为特征具有明显优势。

关键词:静态手势识别;手势交互系统;Hu 矩;支持向量机

中图分类号:TP391.4

文献标识码:A

文章编号:1673-629X(2016)08-0200-05

doi:10.3969/j.issn.1673-629X.2016.08.043

Research and Application of Gesture Recognition Based on Information of Body Skeleton and Depth

WU Cai-fang¹, XIE Jun¹, YU Lu²

(1. School of Command Information System, PLA University of Science and Technology,
Nanjing 210007, China;
2. School of Communication Engineering, PLA University of Science and Technology,
Nanjing 210007, China)

Abstract: It explores the technology of gesture recognition based on Kinect in this paper, and designs and implements a small gesture interactive system with perfect function and excellent performance. At first, the Skeleton Information and Depth Information from Kinect is used to track and extract hand from the background. The result doesn't be affected by the background, light, and experimenter's skin color and costume. Then, according to the distribution of noise from preliminary acquisition in the binary images, a method of filtering the small scale connected component pixels is put forward to denoise. At last, Hu's moments of hand binary images and hand contour binary images are used as features to train the Support Vector Machine (SVM) classifiers respectively. The experimental results show that compared with Hu's moments of hand binary images, the Hu's moments of the hand contour binary images have obvious advantages.

Key words: static gesture recognition; gesture interactive system; Hu's moment; support vector machine

0 引言

手势识别是通过提取手部数据特征,经过处理和分析,使得计算机能从这些数据中识别出手势。手势识别和人脸识别、语音识别、指纹识别等一样,都是使计算机模拟人脑来识别自然环境中事物的技术。鼠标和键盘是目前广泛使用的人操控计算机的设备,然而随着科技的发展,出现了一些新的人机交互场景,传统

的鼠标键盘不能满足它们的需求,手势识别作为更加自然的人机交互方式逐渐被人们研究和应用。计算机通过识别手势触发相应的命令,可以用于大型交互界面的控制、电脑游戏手柄的替代、远程操控机器人进行高精度作业、与 3D 虚拟环境的交互^[1]等。

手势识别通常分为静态手势识别和动态手势识别两种。静态手势识别是对静态的手型进行识别,动态手势识别是根据手的移动轨迹、速度对手所做出的动

收稿日期:2015-09-17

修回日期:2015-12-23

网络出版时间:2016-08-01

基金项目:国家“863”高技术发展计划项目(2012aa01a509, 2012aa01a510)

作者简介:吴彩芳(1991-),女,硕士研究生,研究方向为模式识别;谢 钧,副教授,研究方向为智能信息处理;俞 璐,副教授,研究方向为计算机网络。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20160801.0842.002.html>

作进行识别。文中主要针对静态手势识别进行研究和应用。静态手势识别通常分为手部分割、手部跟踪、特征提取、分类识别几部分。传统的获取数据的方法有基于数据手套和基于计算机视觉信息两种。使用数据手套的方法^[2]虽然精度高,但是设备昂贵,使用舒适感欠佳,难以推广,因此现在一般都采用基于计算机视觉的方法获取数据^[3]。手部分割是将手从背景中分离出来,通常有基于运动的分割方法、基于特殊颜色标记的分割方法、建立肤色模型进行肤色检测法^[4]、深度信息法^[5]等。特征提取可以提取像素特征、统计特征、Hu矩特征、手指个数、指尖夹角、手指名称、手指方向等特征^[5]。分类识别主要有模板匹配法^[6]、支持向量机(SVM)^[7]、神经网络法、隐马尔可夫模型法^[8]等。

文中使用微软推出的 Kinect 体感设备进行手势识别的研究和实现,首次结合使用 Kinect 骨骼信息和深度信息的提取并跟踪手部,使得后续的手势识别操作更加灵活。选取 Hu 矩作为静态手势的特征,Hu 矩因兼具旋转不变性、平移不变性、尺度不变性等特点而被广泛用于手势识别的研究中。选取 SVM 分类器进行手势分类识别。实验对比了分别使用同一静态手势的手型 Hu 矩和手型轮廓 Hu 矩进行识别的效果,并分析原因。最后实现一个功能完善的小型手势交互系统。

1 骨骼信息和深度信息的获取及预处理

Kinect 传感器有两个摄像头和一个投影机,分别是彩色摄像头、深度摄像头和红外投影机。红外投影机发出激光,遇到物体反射回来被深度摄像头接收,生成景深图像,并且追踪和提取出视野范围内人的骨骼信息^[9]。实验所使用的 Kinect2.0 是微软的第二代 Kinect 体感设备,深度摄像头的可视范围是距离摄像头平面 0.5 m 到 8 m 之间的范围,能够检测到的人体骨骼节点有 25 个,其中与实验相关的节点有腕关节点、掌指关节点、食指指尖点和拇指指尖点。深度摄像头每秒生成 30 帧景深图像,景深图像中每个像素点上的数值代表物体与 Kinect 摄像头的距离,如果值为 0,则表示该点处在视野范围内没有发现任何物体。Kinect 追踪并提取的人体骨骼关节点坐标信息中的 Z 坐标值是节点距摄像头的距离。通过 Kinect 获取的人体骨骼关节点信息可以定位到手的位置,根据骨骼关节点之间深度值的关系设定深度阈值,提取出手部像素点。

1.1 手部追踪与提取

进行手势识别时,实验者位于 Kinect 体感设备正前方 0.5 m 到 8 m 之间的范围内,面向摄像头,确保身体前方没有遮挡物,伸出一只手,Kinect 获取实验者伸

出的手的腕关节点、掌指关节点、食指指尖点和拇指指尖点的深度值,取其中的最大值赋值给 max,最小值赋值给 min。将 Kinect 获取到的深度图像通过以下方法转化为二值图像:深度值在阈值范围 $[\min - a, \max]$ 内的像素点的灰度值置为 1,其他像素点的灰度值置为 0。 a 是使该阈值范围刚好囊括所有手部像素点的最小值。实验结果表明, a 的值取 10 可以满足上述要求。该二值图像是一个粗糙的手型图像,大部分的背景点是黑色,大部分的手部像素点是白色,二值图中有一些噪声点待处理。初步获取的手部图像如图 1 左所示。

1.2 二值图去噪

从图 1 可以看出,通过深度阈值提取的手部像素点生成的二值图主要存在的问题就是画面中有些孤立的噪声点。这些误把背景识别成手或是把手误识成背景的像素点比较分散,分布没有规律。由于图片是二值图,通常使用的均值滤波、中值滤波、高斯滤波等方法不适用这种情况。文中提出一种简单实用的方法,可以很好地解决这种带有离散噪声点的二值图的去噪问题。

提取二值图中的连通分量,使像素点个数小于 100 的连通分量被周围像素点多的大连通分量吞并,即像素点个数小于 100 的小连通分量像素点的值置为包围它的大连通分量像素点的值,返回生成的二值图。生成的二值图如图 1 右所示。



图 1 初步提取的效果图和二值图去噪后的效果图

1.3 手型轮廓提取

文献[10]使用手型图像 Hu 矩作为静态手势数据特征,文献[11]使用手型轮廓 Hu 矩作为静态手势数据特征,两篇文献都使用了 SVM 分类器来识别手势。为了探索使识别率最大化的数据特征,文中对同一静态手势分别用手型 Hu 矩和手型轮廓 Hu 矩做数据特征,用 SVM 分类器识别,对比识别效果并分析原因。下面进行手型轮廓提取。

假设集合 C 为手部的轮廓点集。指针 s 指向整幅图片第一个扫描到的手部像素点。

(1) 从上到下,从左到右依次扫描图片像素点,扫描到第一个手部像素点后将指针 s 和 p 指向该点。

(2) 从 p 上方像素点开始顺时针扫描 p 周围的 8 个像素点,直到遇到第一个手部像素点,用指针 q 指向

该点,将 p 指针指向的点加入集合 C 中, p 指针指向 q 指针指向的点。

(3) 循环进行步骤 2, 直至 p 指针指向的点和 s 指针指向的点再次重合。返回集合 C 中的像素点。返回生成的轮廓图如图 2 所示。



图 2 手势轮廓提取

1.4 特征提取

文中选取的数据特征是 Hu 矩。Hu 矩主要表征了图像区域的几何特征,具有旋转、平移、尺度不变性。Hu 矩是由七个矩不变量组成的特征,这七个不变矩是由图片的二阶和三阶的中心矩的线性组合构成的^[12]。对于连续的图像 $f(x, y) (\geq 0)$, $p + q$ 阶几何矩 m_{pq} 定义为:

$$m_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q f(x, y) dx dy$$

$p + q$ 阶中心矩 μ_{pq} 定义为:

$$\mu_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \bar{x})^p (y - \bar{y})^q f(x, y) dx dy$$

其中, p 和 q 为非负整数; \bar{x} 和 \bar{y} 为图像的质心。

$$\bar{x} = \frac{m_{10}}{m_{00}}$$

$$\bar{y} = \frac{m_{01}}{m_{00}}$$

$p + q \leq 10$ 时的 10 个几何矩都具有物理意义。0 阶矩 (m_{00}) 表示物体的质量; 1 阶矩 (m_{10}, m_{01}) 表示物体的质心; 2 阶矩 (m_{20}, m_{02}, m_{11}) 表示旋转半径; 3 阶矩 ($m_{30}, m_{03}, m_{12}, m_{21}$) 描述物体的方位和斜度。显然这些矩都不兼具平移、旋转、缩放不变性。因此, 基于这 10 个矩, Hu. M. K 在 1962 年提出了一组不变矩, 并证明它们具有平移、旋转、缩放不变性。归一化的中心矩定义为:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^\gamma}, \gamma = \frac{p + q + 2}{2}$$

七个 Hu 不变矩的定义如下:

$$\Phi_1 = \eta_{20} + \eta_{02}$$

$$\Phi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2$$

$$\Phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \mu_{03})^2$$

$$\Phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \mu_{03})^2$$

$$\Phi_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{30})[3(\eta_{12} + \eta_{30})^2 - (\eta_{21} + \eta_{03})^2]$$

$$\Phi_6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{12} + \eta_{03})$$

$$\Phi_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] - (\eta_{30} - 3\eta_{12})(\eta_{12} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]$$

2 手势识别

文中采用 SVM 作为手势分类识别的工具。SVM 是一种由统计学发展而来的机器学习方法。主要思想是通过一个非线性映射 p , 把样本空间映射到一个高维乃至无穷维的特征空间中, 使得在原来的样本空间中非线性可分的问题转化为在特征空间中线性可分的问题。对于分类、回归等问题来说, 很可能在低维样本空间无法线性处理的样本集, 在高维特征空间中却可以通过一个线性超平面实现线性划分 (或回归)。一般的升维都会带来计算的复杂化, 而 SVM 应用核函数的展开定理巧妙地解决了这个难题, 就不需要知道非线性映射的显式表达式; 由于是在高维特征空间中建立线性学习机, 所以与线性模型相比, 不但几乎不增加计算的复杂性, 而且在某种程度上避免了“维数灾难”^[13]。

实验采用 LIBSVM 工具包。LIBSVM 是台湾大学林智仁教授等开发设计的一个简单易用、快速有效的 SVM 模式识别与回归的软件包。该软件对 SVM 所涉及的参数调节相对比较少, 提供很多的默认参数, 利用这些默认参数可以解决很多问题, 并提供了交互检验的功能。该软件可以解决 C -SVM、 v -SVM、 ε -SVR 和 v -SVR 等问题, 包括基于一对一算法的多类模式识别问题^[14]。

3 实验

3.1 手势识别实验

实验目的是对图 3 所示的五种静态手势进行分类识别, 并且应用于手势交互系统中。实验开发环境是 Visual Studio 2012, 编程语言使用 C#, 硬件设备使用 Kinect for Windows 2.0。



图 3 需要识别的手势图

为了完成实验, 编写的程序依次完成如下功能:

(1)五种静态手势每种手势让不同的人来做,共做 260 次,使用 Kinect 记录并生成 260 幅手型二值图和 260 幅对应的手型轮廓二值图。

(2)以每种手势的前 130 张手型二值图为训练样本,后 130 张手型二值图为测试样本,分别计算它们的 Hu 矩并按 LIBSVM 软件包要求的格式生成训练文件和测试文件。调用 LIBSVM 软件包训练 SVM 模型,并测试识别率。

(3)用同样的方法使用手型轮廓二值图 Hu 矩训练模型并测试识别率。

3.2 实验结果

使用手型二值图 Hu 矩作为数据特征的总识别率 Accuracy = 65.68%。使用手型轮廓二值图 Hu 矩作为数据特征的总识别率 Accuracy = 99.23%。手型识别率分别如表 1 和表 2 所示。

表 1 手型原图的识别率











手型原图	识别率/%
	46.2
	93
	25.3
	96.2
	67.7

表 2 手型轮廓图片的识别率

手型轮廓	识别率/%
	99.23
	100
	100
	100
	96.92

从上述实验结果可以看出,用手型轮廓二值图 Hu 矩做数据特征的识别率明显比用手型二值图 Hu 矩做数据特征的识别率高。出现这种差异的原因,是不同手势的手型原图中共同点较多,而不同手势的手型轮廓共同点很少。手型原图中手势用手部的全部像素点表示,而手型轮廓图中手势只用手型的一圈外轮廓像素点表示。在手型原图中,每种手势的手掌部分面积占整个手型面积的一半以上,而手掌部分像素点是每种手势的共同点;手型轮廓图只保留了手型的外轮廓像素点,去掉了不同手势相同的像素点,拉开了不同手势之间的差异,因此用手型轮廓图表示静态手势更具有代表性,使用手型轮廓 Hu 矩作为静态手势的数据特征进行识别的识别率更高。

接下来将使用手势轮廓 Hu 矩训练好的 SVM 模型,分类识别这五种手势,据此实现一个功能完善的小型手势交互系统。

3.3 手势交互系统的实现

从上述实验结果可以看出,已经做到了对上述五种手势的分类识别。而且,调用 Kinect 体感设备的内置接口可以实现手势的握拳和松拳的识别。利用这些条件,设计出一个手势交互系统。该系统实现了鼠标的左键和中键的功能,实现了利用手势远程控制计算机上音乐播放器、视频播放器、IE 浏览器等常用的应用。与计算机交互时,操作人员只需站在 Kinect 视野范围内,面向 Kinect 摄像头,伸出手做上述的几种手势,即可远程操控计算机上的应用。若做出某一手势前计算机上相应的应用还未打开,则计算机做出回应,打开该应用;否则,计算机忽视该手势命令。系统的工作流程图如图 4 所示。

分别让 10 个人在 10 个不同的场景中测试该系统,每人每个手势操作 10 次,记录用手势操控计算机的出错率,结果如表 3 所示。

表 3 手势交互系统使用测试

手势命令名称	出错率/%
鼠标移动	0
鼠标左键单击	0
鼠标滑轮上滚	2
鼠标滑轮下滚	1
打开视频播放器	2
打开音乐播放器	1
打开 IE 浏览器	4

通过测试结果可以看出,该手势交互系统的鲁棒性强,识别率与背景、光线等因素无关。使用手势控制对计算机进行操作出错次数少,出错成本低,用户使用的舒适度较好。用户操作的空间范围大,使用灵活。

4 结束语

文中提出一种结合 Kinect 获取的骨骼信息和深度信息进行手势分割和跟踪的方法,以手势轮廓 Hu 矩为数据特征,采用 SVM 分类器对五类静态手势进行分类识别,并且以该方法实现了一个功能完善的小型手势交互系统。实验结果表明,该方法识别率高、灵活性强,手势交互系统鲁棒性强、简单实用。

虽然文中实现的手势识别方法精度高,但是识别的手势种类少,设计的手势交互系统简单,因此下一步将结合动态手势识别,研究更复杂的手势识别,实现功能更加丰富的手势交互系统。Kinect 体感设备提供了丰富的功能接口,手势识别只使用了其中很少的一部

分。除了可以识别人体骨骼, Kinect 还提供了人脸识别、人眼跟踪、语音识别、表情识别等功能接口, 下一步

可以结合这些接口开发出更加自然的人机交互系统。

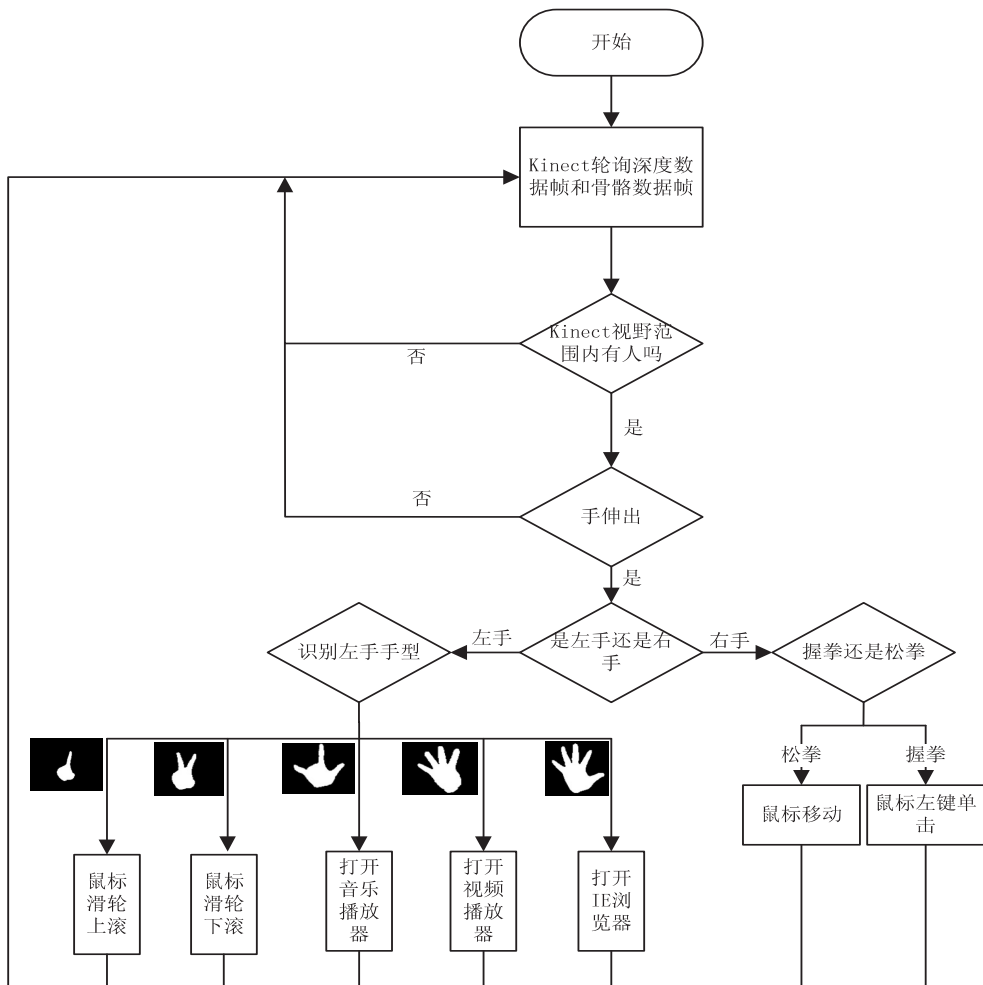


图 4 手势识别交互软件的流程图

参考文献:

[1] 翟言. 体感交互技术在大型沉浸式系统中的应用与研究[D]. 北京:北京林业大学,2014.

[2] 吴江琴,高文,陈熙霖. 基于数据手套输入的汉语手指字母的识别[J]. 模式识别与人工智能,1999,12(1):74-78.

[3] Mitra S, Acharya T. Gesture recognition: a survey[J]. IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications and Reviews,2007,37(3):311-324.

[4] 廖赞. 基于裸手的自然人机交互关键算法研究[D]. 昆明:云南大学,2012.

[5] Li Yi. Hand gesture recognition using Kinect[C]//Proceedings of 2012 IEEE 3rd international conference on software engineering and service science. [s. l.]: IEEE,2012:196-199.

[6] 李文杰. 基于骨架化和模板匹配的交通指挥手势识别[D]. 杭州:浙江工业大学,2011.

[7] 李云峰,胡文平. SVM 多类分类算法及其在手写体数字识别中的应用[J]. 物流工程与管理,2012(7):131-134.

[8] 洪文,黄凤岗,苏茜. 基于连续隐马尔科夫模型的步态

识别[J]. 应用科技,2005,32(2):50-52.

[9] 余淘. Kinect 应用开发实战:用最自然的方式与机器对话[M]. 北京:机械工业出版社,2012.

[10] 刘云,甘志杰. 基于 Hu 矩和支持向量机手完整性检测算法及应用[J]. 计算机工程与应用,2008,44(30):223-225.

[11] Liu Yun, Yin Yanmin, Zhang Shujun. Hand gestures recognition based on Hu moments in interaction of virtual reality[C]//Proc of the intelligent human-machine systems and cybernetics. [s. l.]:[s. n.],2012:11-14.

[12] Huang Zhihu, Lengng Jinsong. Analysis of Hu's moment invariants on image scaling and rotation[C]//Proc of 2010 2nd international conference on computer engineering and technology. [s. l.]:[s. n.],2012:476-478.

[13] Chang Chih-Chung, Hsu Chih-Wei, Lin Chih-Jen. The analysis of decomposition methods for support vector machines[J]. IEEE Transactions on Neural Networks,2000,11(4):1003-1008.

[14] Chang Chih-Chung, Lin Chih-Jen. LIBSVM: a library for support vector machines[J]. ACM Transactions on Intelligent Systems and Technology,2011,3(2):23-30.